5Я

УДК 621.39

І.М. Яремко, В.В. Турупалов
Донецький національний технічний університет, Україна
son@dn.farlep.net

Управління розподілом ресурсів центрів обробки даних телекомунікаційної мережі

Розглянуто проблему оптимізації ресурсів центрів обробки даних телекомунікаційної мережі. Запропоновано комбінований алгоритм багатокритеріальної оптимізації методу послідовних поступок та методу обмежень, який забезпечує розв'язання задачі оптимізації кількості серверів центрів обробки даних за кількома критеріями.

Вступ

Розвиток інформаційних технологій пов'язаний із реалізацією концепції центрів обробки даних (ЦОД) – комплексних організаційно-технічних рішень для створення високопродуктивної, відмовостійкої ІТ-інфраструктури. До головних завдань ЦОД належать консолідоване зберігання і опрацювання даних користувачів, надання їм прикладних сервісів, підтримка функціонування застосувань.

В умовах глобальної інформатизації суспільства і бізнесу зростають вимоги користувачів до рівня сервісу, керованості, надійності, доступності і масштабованості ІТ-інфраструктури [1]. Це ускладнює управління ІТ-інфраструктурою. Створення ЦОД вимагає значних коштів, ефективної підтримки ІТ-інфраструктури, наявності у штаті висококваліфікованих фахівців. Тому все більше фірм стають клієнтами хостингових компаній.

Постановка проблеми

Проблема створення і забезпечення ефективного функціонування ЦОД постає насамперед перед хостинговими компаніями. Вкладаючи кошти, компанії сподіваються на прибуток. У будь-якому випадку вони очікують зменшення витрат на експлуатацію ЦОД, зниження вартості обслуговування користувачів, що дозволить, зрештою, закласти основу для ефективної діяльності як самої компанії, так і клієнтів.

Клієнти своє бачення роботи ІТ-інфраструктури погоджують із хостинговою компанією на рівні вимог, до яких звичайно належать: вартість послуг; доступність і керованість ІТ-інфраструктури; цілісність даних; безпека; надійність; масштабованість. Досягнення рівня вимог користувачів найменшими коштами становить сутність проблеми створення і забезпечення функціонування ЦОД. Зазвичай цю комплексну проблему розбивають на ряд проблем менших розмірів, але від того не набагато простіших. Однією з них є проблема управління ресурсами і навантаженням ЦОД.

Аналіз і розв'язання проблеми

Середній час відповіді на запит і ймовірність перевищення часу відповіді від заданої величини є одними з найбільш важливих показників, що характеризують

якість роботи хостингової компанії. Вона може вплинути на ці показники як на рівні центру обробки даних (вузла), так і мережі в цілому. На рівні вузла це реалізується шляхом включення додаткових серверів у кластери. Однак число серверів повинне визначатися залежно від вхідного навантаження, інакше устаткування провайдера буде недостатньо завантажене й, як наслідок, провайдер буде нести додаткові витрати. Тому знаходження оптимального числа серверів у центрі обробки даних є актуальним завданням.

Завдання оптимізації ґрунтуються на розробленій моделі центру обробки даних, що має триланкову архітектуру (Web-сервери, сервери застосувань і сервери баз даних) і що є найпоширенішою. Для розв'язання цього завдання поставлено завдання оптимізації ресурсів центрів обробки даних із двома критеріями (витрати й пропускна здатність) і обмеженнями на середній час відповіді й імовірність перевищення часу відповіді від заданої величини й запропонована технологія їх розв'язання.

Постановка двокритеріальної задачі оптимізації для визначення оптимальної кількості серверів без урахування класів запитів за критеріями сукупної вартості володіння серверами ЦОД і пропускної здатності:

$$\max_{\{S_1,\dots,S_C\}} F_{ПС}(S_1,\dots,S_C) = \sum_{c=1}^{C} \frac{S_c}{m_{Tc}},$$

$$\min_{\{S_1,\dots,S_C\}} F_{TCO}(S_1,\dots,S_c) = \sum_{c-1}^{C} TCO_c \cdot S_c,$$

при обмеженнях

$$\sum_{c=1}^{C} \frac{\lambda_c}{\lambda} \frac{m_{tC}}{S_c - q_c} \le t^*_{SLA}, \quad q_c < S_c \le S_c^{max}, \; c = \overline{1,C}$$

де C – кількість кластерів у ЦОД; S_c – число серверів у кластері c; $F_{ПС}(S_1,\dots, S_C)$ – критерій пропускної здатності; $F_{TCO}(S_1,\dots, S_C)$ – критерій сукупної вартості володіння серверами ЦОД; TCO_c – сукупна вартість володіння одним сервером кластера; $\frac{\lambda_c}{\lambda}$ – середня кількість відвідувань кластера c запитом за час його перебування в системі; q_c – номінальне завантаження кластера c з одним сервером при заданому навантаженні λ_c; S_c^{max} – максимально можлива кількість серверів у кластері c; m^{Tc} – середній час відповіді на запит сервером кластера c; t^*_{SLA} – оптимальний середній час відповіді, визначений з аналізу SLA.

Постановка задачі для обмеження на середній час відповіді на запит і / або максимальний час відповіді на запит методики оптимізації знаходимо оптимальну кількість серверів, для заданої долі всіх запитів з поділенням цих запитів на класи:

$$\min_{\{S_1,\dots,S_C\}} F_{TCO}(S_1,\dots,S_c) = \sum_{c=1}^{C} TCO_c \cdot S_c,$$

$$\max_{\{S_1,\dots,S_C\}} F_{ПЗ}(S_1,\dots,S_C) = \sum_{c=1}^{C} \frac{S_c}{\frac{1}{K}\sum_{k=1}^{K} m_c^k},$$

при обмеженнях

$$\sum_{c=1}^{C} \frac{\lambda^k{}_c}{\lambda^k} \frac{m^k{}_c}{S_c - q_c} \le T^k{}_{SLA}, \; k = \overline{1,K} \quad q_c < S_c \le S_c^{max}, \; c = \overline{1,C},$$

де C – число кластерів ЦОД; S_c – число серверів у кластері c; $F_{ПС}(S_1, \ldots, S_C)$ – критерій пропускної здатності; $F_{TCO}(S_1, \ldots, S_C)$ – критерій сукупної вартості володіння; TCO_c – сукупна вартість володіння одним сервером кластера c; q_c – номінальне завантаження кластера c з одним сервером при заданому навантаженні; m_c^k – середній час відповіді на запит класу k сервером кластера с; $\frac{\lambda^k{}_c}{\lambda^k}$ – середня кількість відвідувань кластера c запитом класу k за час його перебування в системі; T_{SLA}^k – середній час часу відповіді на запит класу k, яке задається в SLA-угоді; S_c^{max} – максимально можлива кількість серверів у кластері c.

Метод обмежень пов'язаний з поняттям компромісного розв'язку й дозволяє знайти ефективний розв'язок точно відповідно до переваги, що задається на безлічі цільових функцій за допомогою вагових коефіцієнтів ρ_i, які визначають відносну важливість i-го критерію порівняно з іншими. Під компромісним розв'язком розуміється існування на безлічі X такого розв'язку, при якому величина відхилень від оптимальних значень за кожним критерієм досягає найменшого значення:

$$\Delta f_i(x) = \begin{cases} f_i^0 - f_i(x) & \forall i \in I_1 \\ f_i(x) - f_i^0 & \forall i \in I_2 \end{cases},$$

де f_i^0 – оптимальне значення i-ї функції мети на безлічі X; I_1 – безліч максимізованих критеріїв; I_2 – безліч минімізованих критеріїв.

Оскільки найменше значення величини $\Delta f_i(x)$ не досягається одночасно за всіма критеріями при жодному розв'язку, то виникає необхідність порівняння цих величин між собою, що вимагає залучення в ситуацію ухвалення складного рішення додаткової інформації від експертів. При цьому необхідно визначити кількісні характеристики для порівняння величин відхилень від оптимальних значень критеріїв різної розмірності між собою й переваг критеріїв на безлічі цільових функцій, з урахуванням яких ухвалюється розв'язок. Для цього вводять монотонні перетворення $w_i(f_i(x))$, які приводять окремі критерії $f_i(x)$ до безрозмірного виду й дозволяють порівнювати їх між собою. Після таких перетворень приватних критеріїв можна одержати зважену відносну величину відхилень від оптимальних значень по кожній цільовій функції $\rho_i w_i(f_i(x))$, де ρ_i – вагові коефіцієнти окремих критеріїв, які ще називають коефіцієнтами переваги окремих критеріїв.

На основі теореми 1.3.1. [2] компромісний розв'язок $x^* \in X$ є розв'язком завдання векторної оптимізації, якщо воно забезпечує однакові мінімальні зважені відносні втрати (відхилення) $\rho_i w_i(f_i(x))$ за всіма критеріями одночасно.

Процес розв'язання завдання методом обмежень складається із двох етапів. Перший етап полягає в завданні вектора коефіцієнтів переваги приватних критеріїв ρ і відшуканні однакових мінімальних зважених відносних k_0^{min} втрат критеріїв шляхом побудови ітераційного процесу з параметром $k_0 \in (0{,}1/M)$, де M – число частинних критеріїв, на кожному кроці якого перевіряється спільність системи нерівностей:

$$\rho_i \omega_i(x) \le k_0$$

для $x \in X$ й заданого вектора ρ. Параметр k_0 обмежує відносні зважені втрати $\rho_i \omega_i(x) \quad \forall i \in I_0$

Якщо при знаходженні k_0^{min} знайдений розв'язок є єдиним, то він є й компромісним розв'язком. Якщо він не єдиний, то виконується другий етап процесу розв'язання завдань векторної оптимізації. Другий етап полягає у виборі й мінімізації узагальненого критерію виду:

$$F(x) = \sum_{i=1}^{M} \rho_i \omega_i(x).$$

На безлічі припустимих розв'язків $X' = \{x \mid x \in X, \rho_i \omega_i(x) \le k_0^{min}, \forall i = \overline{1,M}\}$. У результаті його виконання одержуємо компромісний розв'язок.

При використанні даного методу як монотонні перетворення окремих критеріїв *f(x)* вибираються співвідношення:

$$\omega_i(x) = \begin{cases} \dfrac{f_i^0 - f_i(x)}{f_i^0 - f_i^{\min}} & \forall i \in I_1 \\ \dfrac{f_i(x) - f_i^0}{f_i^{\max} - f_i^0} & \forall i \in I_2 \end{cases}.$$

Перевага методу обмежень полягає в тому, що він слушний для будь-якого виду функцій цілі $f_i(x)$ й обмежень. Недоліком методу є те, що при знаходженні k_0^{min} не завжди знайдений розв'язок є єдиним, що викликає необхідність формування узагальненого критерію $F(x)$, і як наслідок, появу його недоліків.

Метод послідовних поступок застосовується, коли на основі приватних критеріїв можна проводити якісний аналіз відносної важливості цих критеріїв. На основі такого аналізу критерії розташовуються й нумеруються в порядку убування важливості.

Нехай приватні критерії розташовані в порядку убування важливості $Z_1(x) \succ Z_2(x) \succ \cdots \succ Z_M(x)$. Припустимо, що всі вони максимізуються, тоді компромісним розв'язком уважається розв'язок останньої підзадачі з наступної послідовності завдань.

$$Z_1^{max} = \max_{x \in X} Z_1(x) \tag{1}$$

$$\begin{gathered} Z_2^{компр} = \max_{x \in X} Z_2(x) \\ Z_1(x) \ge Z_1^{max} - \Delta_1^{компр} \end{gathered} \tag{2}$$

$$\begin{gathered} Z_j^{компр} = \max_{x \in X} Z_j(x) \\ Z_1(x) \ge Z_1^{max} - \Delta_1^{компр} \\ \cdots\cdots\cdots \\ Z_{j-1}(x) \ge Z_{j-1}^{max} - \Delta_{j-1}^{компр} \end{gathered} \tag{3}$$

$$\begin{gathered} Z_M^{компр} = \max_{x \in X} Z_M(x) \\ Z_r(x) = Z_r^{компр} - \Delta_r, \quad r = \overline{1, M-1} \end{gathered} \tag{4}$$

Сутність методу полягає в розв’язанні послідовності підзадач (1-4), починаючи з першої, й послідовному визначенні величин поступок $\Delta^1, \Delta^2, \Delta^3, \ldots$. Процедура визначення необхідної величини поступки $\Delta^{компр}$ складається з багаторазового повторення розв’язку кожної підзадачі шляхом завдання пробних значень поступок і аналізу залежності від них результатів розв’язку. Компромісним розв’язком буде розв’язок останньої підзадачі.

Перевагою методу послідовних поступок є те, що він може бути використаний для будь-яких частинних критеріїв і обмежень, що мають різну розмірність. Даний метод не вимагає монотонних перетворень частинних, тому що він не вимагає порівняння їх значень. Крім цього, він простий у використанні й наочний.

Довільний вибір величин пробних поступок, число яких може бути більшим, призводить у ряді завдань до значних часових витрат, що є основним недоліком даного методу, але є мінімальним при розв’язанні двокритеріальних задач оптимізації. Тому з урахуванням даної обставини й перерахованих переваг методу доцільно його застосування для розв’язання поставленого завдання оптимізації. З метою одноетапного знаходження величини поступки можна запропонувати процедуру, засновану на комбінації методів послідовних поступків і обмежень.

Сутність пропонованого комбінованого алгоритму методу послідовних поступок полягає в завданні експертом ОПР кількісних оцінок переваги частинних критеріїв, одноетапному визначенні величини компромісної поступки $\Delta_1^{компр}$ й наступному розв’язання підзадач (1-4) з метою знаходження розв’язку поставленого завдання.

Одноетапність визначення величини компромісної поступки дозволяє значно скоротити час розв’язання завдань оптимізації розподілу ресурсів ЦОД і зменшити число звернень до ОПР.

Процес розв’язання завдання оптимізації розподілу ресурсів ЦОД модифікованим методом послідовних поступків складається з наступних етапів.

1. На першому етапі задається вектор переваг частинних критеріїв:

$$\rho = \left\{ \rho_i : \rho_i > 0 \forall i = \overline{1, M}, \sum_{i=1}^{M} \rho_i \right\}$$

де ρ_i – коефіцієнт переваги i-го критерію ($i = \overline{1, M}$). Проводиться впорядкування критеріїв, тобто критерію з більшим коефіцієнтом переваги привласнюється перший номер.

2. Другим етапом процесу розв’язання завдання є розв’язання наступної послідовності підзадач: знайти максимальні й мінімальні значення для кожного критерію без урахування інших критеріїв.

Знайти мінімальні зважені відносні втрати (відхилення) окремих критеріїв від своїх оптимальних значень k_0^{min} [2]. Для цього перетворимо критерії до безрозмірного виду, використовуючи монотонно-убутні функції $\omega(x)$:

$$\omega_i(x) = \omega_i(Z_i(x)) = \frac{Z_i^{max} - Z_i(x)}{Z_i^{m\,max} - Z_i^{min}}, \qquad i = \overline{1, M}.$$

Згідно з методом обмежень, k_0^{min} знаходиться розв’язанням системи нерівностей

$$\rho_i \omega_i(x) \le k_0 \qquad i = \overline{1, M}$$

для мінімального параметра k_0, при якому система сумісна.

$$\Delta_i^{компр} = \frac{k_0^{min}}{\rho_i}\left(Z_i^{max} - Z_i^{max}\right), i = \overline{2,M} \ .$$

Тоді компромісний розв’язок завдання є розв’язком останнього завдання з наступної послідовності завдань (1-4).

Оптимальний розв’язок завдання розподілу ресурсів ЦОД і відповідне йому значення критеріїв – результат розв’язання останнього завдання (4).

Висновки

Процедура одноетапного визначення компромісного розв’язку, заснована на комбінації методів послідовних поступків і обмежень, дозволяє значно скоротити час розв’язання завдання, особливо якщо критеріїв більше двох, а це, у свою чергу, дозволить провайдеру включити алгоритм багатокритеріальної оптимізації в контур управління ресурсами центрів обробки даних телекомунікаційної мережі.

Література

1. Вентцель Е.С. Введение в исследование операций / Е.С. Вентцель. – М. : Советское радио, 1964.
2. Михалевич В.С. Вычислительные методы исследования и проектирования сложных систем / В.С. Михалевич, В.Л. Волкович. – М. : Наука, 1982. – 286 с.
3. Подиновский В.В. Оптимизация по последовательно применяемым критериям / В.В. Подиновский, В.М. Гаврилов. – М. : Советское радио, 1975. – 192 с.

Literatura

1. Ventcel’ E.S. Vvedenie v issledovanie operacij. M.: Sovetskoe radio. 1964.
2. Mihalevich V.S. Vychislitel’nye metody issledovanija i proektirovanija slozhnyh sistem. M.: Nauka. 1982. 286 s.
3. Podinovskij V.V. Optimizacija po posledovatel’no primenjaemym kriterijam. M.: Sovetskoe radio. 1975. 192 s.

И.Н. Яремко, В.В. Турупалов

Управление распределением ресурсов центров обработки данных телекомуникационной сети
Рассмотрена проблема оптимизации ресурсов центров обработки данных телекоммуникационной сети. Предложен комбинированный алгоритм многокритериальной оптимизации метода последовательных уступок и метода ограничений, который обеспечивает решение задачи оптимизации количества серверов центров обработки данных по нескольким критериям.

I.N. Yaremko,V.V. Turupalov

Management of Resources for Telecommunication Network Data Center
The problem of optimizing resources for telecommunication networkdata center is analized. We propose a combined multi-objective optimization algorithm of successive concessions and limitation method, which provides solution for the problem of optimizing the number of servers in data centers according to different criteria.

Стаття надійшла до редакції 08.07.2011.